Comments on "Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach" by G. Doppelhofer, R. Miller, and X. Sala-i-Martin

Manuel Arellano

Barcelona, June 5, 2000

Introduction

These are the comments of an outsider!

Barro (1991) originated an explosion of empirical research on the determinants of growth using cross-country data. Once this became a "mature" literature, people started to ask what had been learned from it.

1) Some thought that explaining growth rates by differences in initial levels was not particularly interesting after all, and sought a shift of focus towards the determinants of *differences in levels* of activity across countries (as Hall & Jones, 1999). Others went in the opposite direction by resorting to *panel data*, trying to obtain better measures of the dynamics of growth rates.

2) A more specific issue was raised by Levine & Renelt (1992). From an *extreme-bounds analysis,* they concluded that no variable was robustly correlated with growth. In contrast, Sala-i-Martin (1997) considered weighted-averages of OLS coefficients and found that some were fairly stable.

The present paper is concerned with the answer to the question “what can be learned ...” in the second, narrower context.

The *contributions of the paper* are:

- 1) Providing a statistical grounding for LR weighted averages of OLS coefficients.
- 2) Generalizing the argument to regressions with different numbers of coefficients.
- 3) Showing there is a subset of variables strongly related to growth.

The paper is well written an extremely clear. So I will not try an alternative description of what the paper does.

Summary of the approach

The paper considers the posterior distribution of β as

$$g(\beta \mid y) = \sum_j \Pr(M_j \mid y)\, g(\beta \mid y, M_j).$$

M_j denotes a model (a list of included variables). Given M_j, $g(\beta \mid y, M_j)$ is standard in a normal regression with diffuse priors (the sampling distribution of OLS).

The *odd posterior ratio* between two models is given by the *odd prior ratio* times the *likelihood ratio LR*:

$$\frac{\Pr(M_j \mid y)}{\Pr(M_\ell \mid y)} = \left(\frac{\Pr(M_j)}{\Pr(M_\ell)}\right) \times \left(\frac{f(y \mid M_j)}{f(y \mid M_\ell)}\right)$$

They consider the Schwarz *adjusted LR* which is the *unadjusted LR* times a correction factor for the difference in the no. of parameters

$$\frac{f(y \mid M_j)}{f(y \mid M_\ell)} = T^{-(k_j - k_\ell)/2} \left(\frac{\widehat{\sigma}_j^2}{\widehat{\sigma}_\ell^2}\right)^{-T/2}.$$

Moreover, they specify the odd prior ratio as

$$\frac{\Pr(M_j)}{\Pr(M_\ell)} = \left(\frac{\overline{k}}{K}\right)^{k_j - k_\ell} \left(1 - \frac{\overline{k}}{K}\right)^{k_\ell - k_j}$$

where $\overline{k}$ is a prior about the *no. of regressors with non-zero coefficients*, and K $(= 32)$ is the total no. of regressors.

So two models with the same no. of parameters have the same prior probabilities, and any model with $\overline{k}$ regressors has the largest prior probability, regardless of the nature of the regressors.

Reported statistics:

1) Posterior mean:

$$E\left(\beta_i \mid y\right) = \sum\nolimits_j \Pr\left(M_j \mid y\right) E\left(\beta_i \mid y, M_j\right).$$

2) Posterior variance:

$$Var\left(\beta_i \mid y\right) = E\left[Var\left(\beta_i \mid y, M_j\right) \mid y\right] + Var\left[E\left(\beta_i \mid y, M_j\right) \mid y\right.$$

3) Posterior inclusion probability:

$$\Pr\left(\beta_i \neq 0 \mid y\right) = \sum\nolimits_{\beta_i \neq 0} \Pr\left(M_j \mid y\right).$$

4) Sign certainty probability:

$$\Pr\left[sgn\left(\beta_i\right) = sgnE\left(\beta_i \mid y\right) \mid y, \beta_i \neq 0\right]$$

5) Posterior means and variances conditional on inclusion.

6) Graphics of $g\left(\beta_i \mid y, \beta_i \neq 0\right)$ *and* $\left[1 - \Pr\left(\beta_i \neq 0 \mid y\right)\right]$

In the normal regression with diffuse priors $E\left(\beta_i \mid y, M_j\right)$ coincides with OLS, and $Var\left(\beta_i \mid y, M_j\right)$ with the OLS sampling variance.

$\Pr\left(\beta_i \neq 0 \mid y\right)$ is used as a *ranking* measure of how much the data favors the inclusion of a variable in the growth regression. Variables are divided according to whether

$$\Pr\left(\beta_i \neq 0 \mid y\right) \gtrless \Pr\left(\beta_i \neq 0\right) = \frac{\overline{k}}{K} \left(\simeq 0.22 \text{ with } \overline{k} = 7\right).$$

There are 12 variables for which the data provide support in this sense. Of those, there are 4 for which

$$\Pr\left(\beta_i \neq 0 \mid y\right) \geq 0.95.$$

Finally, a variable is called “robust” if it has a small posterior standard deviation conditional on inclusion and a high sign-certainty probability.

Comments in a Bayesian spirit

On the priors. The prior chosen for $\Pr(M_j)$ is not very meaningful: We assign the same probability to a model as long as it has the same number of variables, no matter what they are.

Sometimes we are interested not in single variables, but in *groups of variables*. This could be reflected in the choice of priors by assigning higher probabilities to "coherent" specifications.

A Bayesian would be concerned about sensitivity of results to alternative priors of this kind. We may expect the stability of individual coefficients to be affected by the inclusion of related variables.

The choice of priors is a strange one. A variable either enters or does not enter the regression, but if it enters we have diffuse priors about it. This is a "classical prior".

Conditioning on initial income seems to be in the nature of the approach. So why not reducing the number of models by always including initial income?

On the analysis of the posterior distribution.

Even if we retain the priors used in the paper, it may still be interesting to analyze *joint* posterior inclusion probabilities for sets of variables.

Also why focusing on the event $\beta_i \neq 0$ as opposed to probabilities of the β_i taking values within some economically meaningful range?

More generally, in a Bayesian spirit it would be nice to make greater use of the model. For example, by calculating *posterior probability forecasts*.

Let $\varphi(y^*)$ denote some statement about out-of-sample growth. We could calculate probability forecasts of the form

$$\Pr\left(\varphi(y^*) \mid y, x = \overline{x}\right) =$$
$$\sum_j \Pr\left(\varphi(y^*) \mid y, x = \overline{x}, M_j\right) \Pr\left(M_j \mid y, x = \overline{x}\right)$$

where

$$\Pr\left(\varphi(y^*) \mid y, x = \overline{x}, M_j\right)$$
$$= \int \Pr\left(\varphi(y^*) \mid y, x = \overline{x}, M_j, \beta\right) g\left(\beta \mid y, x = \overline{x}, M_j\right) d\beta.$$

Comments in a classical spirit

What *parameter* is being estimated by the posterior mean?

If we regard our 98 countries as a representative sample of a hypothetical population of countries, the posterior mean is a consistent estimate of a weighted average of partial regression coefficients based on different conditioning sets:

$$\overline{\beta} = \sum\nolimits_{j=1}^{2^K} \frac{w_j}{\sum_\ell w_\ell} \beta_j$$

where the weights are

$$w_j = \left(\frac{\overline{k}}{K}\right)^{k_j} \left(1 - \frac{\overline{k}}{K}\right)^{K-k_j} T^{-k_j/2} \left(\sigma_j^2\right)^{-T/2}.$$

Thus a particular partial regression coefficient receives a larger weight the smaller σ_j^2, the smaller k_j, and the closer k_j to $\overline{k}$.

Similarly, the posterior variance matrix is a consistent estimate of

$$\Sigma_\beta = \sum\nolimits_{j=1}^{2^K} \frac{w_j}{\sum_\ell w_\ell} \left[Var\left(\widehat{\beta}_{jOLS}\right) + \left(\beta_j - \overline{\beta}\right)^2\right].$$

It is unclear whether $\overline{\beta}$ is the right measure of the marginal effect of the variable. If there is a "right model", $\overline{\beta}$ will not coincide with the parameters in such model. If there is a subset of "approximately right" models, the discrepancy of $\overline{\beta}$ relative to the coefficients in those models will depend on the covariances between the "right" and "wrong" regressors in ways that are difficult to assess.

Inference

Now let us leave these concerns aside. Let us take for granted an interest in inference about $\overline{\beta}$.

The posterior mean and variance are consistent estimates of $\overline{\beta}$ and Σ_β:

$$\widehat{\overline{\beta}} = \sum_{j=1}^{2^K} \frac{\widehat{w}_j}{\sum_\ell \widehat{w}_\ell} \widehat{\beta}_{jOLS}$$

$$\widehat{\Sigma}_\beta = \sum\nolimits_{j=1}^{2^K} \frac{\widehat{w}_j}{\sum_\ell \widehat{w}_\ell} \left[\widehat{Var} \left(\widehat{\beta}_{jOLS} \right) + \left(\widehat{\beta}_{jOLS} - \widehat{\overline{\beta}} \right)^2 \right]$$

where $\widehat{w}_j$ is similar to w_j after replacing σ_j^2 by its sample counterpart $\widehat{\sigma}_j^2$.

However, if we want a *classic confidence interval* for $\overline{\beta}$, or we wish to perform a test of

$$H_0 : \ \overline{\beta} = 0,$$

$\widehat{\Sigma}_\beta$ is *not* an appropriate measure of the uncertainty in $\widehat{\overline{\beta}}$ as an estimate of $\overline{\beta}$.

From the point of view of inferences about $\overline{\beta}$ there is no data-mining involved in constructing an average of all OLS estimates. $\widehat{\overline{\beta}}$ is just a continuous function $\overline{\beta}(\widehat{\theta})$ of $\widehat{\theta} = (\widehat{\theta}_1, ..., \widehat{\theta}_{2^K})$ where $\widehat{\theta}_j = \left(\widehat{\sigma}_j^2, \widehat{\beta}_{jOLS}\right)$, so that its large sample variance can be approximated by the delta method in the standard way

$$\widehat{Var}\left(\widehat{\overline{\beta}}\right) = \left(\frac{\partial\overline{\beta}(\widehat{\theta})}{\partial\theta'}\right)' \widehat{Var}\left(\widehat{\theta}\right)\left(\frac{\partial\overline{\beta}(\widehat{\theta})}{\partial\theta'}\right).$$

The difference between $\widehat{Var}\left(\widehat{\overline{\beta}}\right)$ and $\widehat{\Sigma}_\beta$ arises for two different reasons:

- Firstly, $\widehat{Var}\left(\widehat{\overline{\beta}}\right)$ takes into account the *sampling correlation* between OLS estimates of different regressions while the posterior variance does not.
- Secondly, $\widehat{Var}\left(\widehat{\overline{\beta}}\right)$ does not take into account population variability in the β_j while $\widehat{\Sigma}_\beta$ does.

Variability in the β_j will depend on the degree of correlation among the various regressors (if they are all orthogonal, all the partial regression coefficients will coincide with the simple regression coefficients).

Comments in a growth empiricist spirit

There is always a tension between the use of *tacit knowledge* and *formal algorithmic methods* in empirical research, because there is always more knowledge about the problem than can be processed in an algorithmic way (as pointed out by Heckman in his 20th Century Retrospective).

This paper belongs to a battle of algorithmic methods. It presents a method to claim robust effects in growth regressions in opposition to another method that claimed lack of robustness. The enterprise is worthwhile. At the very least, to exhibit the lack of robustness of algorithmic searches for robustness.

More importantly, the exercise is very interesting from a forecasting point of view, both methodologically and in terms of the substantive conclusions for forecasts of growth.

But if the objective of growth empirical work is to tell a *convincing story* combining theory and all forms of evidence, I confess to be more sensitive to issues of

- reverse-causality on some of the explanatory variables (like years of openness, the war dummy, or even the rate of population growth),
- confounding convergence rates with unobserved heterogeneity,

than to instability of regression coefficients.

Different researchers have different priors about which variables are important. After all, having a large empirical literature facilitates tailored made combinations of priors, regression results, and posterior conclusions.